REMARKS

Claims 1-20 are in the application.

Claims 17-20 are new.

Claims 1-8, 10-11 and 14-1 6, are rejected under 35 U.S.C. § 102(e) as being anticipated by Paragios et al (U.S. patent 7,139,409). Claim 13 is rejected under 35 U.S.C. § 103(a) as being obvious over Paragios et al. in view of Christopher Jaynes (Dynamic Shadow Removal from Front Projection Displays). Paragios et al. allegedly show a method for detecting shadow regions in an image (column 9, lines 53-57). In fact, the reference appears to indicate that the shadow is an artifact to be removed or ignored from a video sequence which cannot be a single image, since the image sequence is statistically processed using averages over a series of frames from a real-time stream. This is seen, for example, through the use of the words "changes", "arrival", "change detection module", "sensor noise variance", and "real time". Further, the Laplacian transforms *R(s), G(s), B(s)*, are all time based constructs:

II. c) Invariant Normalized Color Module

> Although the color module captures the background intensity properties, it is very sensitive to global illumination changes (e.g. the arrival of a train affects the observed intensities of the platform next to the train line) as well as shadows.
>
> To deal with these limitations introduced by the color based change detection module, a normalization of the RGB color space is preferably performed. As a result, the background properties are not determined by their actual observed values but rather from their relative values in comparison with an associated statistical model.
>
> For example, let (R(s), G(s), B(s)) be the observed color vector. A shadow invariant representation is used, which is given by:

$$\begin{bmatrix} \hat{r}(s) \\ \hat{g}(s) \end{bmatrix} = \begin{bmatrix} \dfrac{\hat{R}(s)}{\hat{R}(s) + \hat{G}(s) + \hat{B}(s)} \\ \dfrac{\hat{G}(s)}{\hat{R}(s) + \hat{G}(s) + \hat{B}(s)} \end{bmatrix}$$

The uncertainties of the

$$\begin{bmatrix} \hat{r}(s) \\ \hat{g}(s) \end{bmatrix}$$

are dependent on the sensor noise variance as well as from the their true values S(s)=(R(s), G(s), B(s)) (due to the non-linearity of the selected transformation). The observed distribution of samples can be approximated using a pixel-wise Gaussian multi-variate distribution given by:

$$\begin{bmatrix} \hat{r}(s) \\ \hat{g}(s) \end{bmatrix} = \begin{bmatrix} \dfrac{R(s) + \eta_R(s)}{S(s) + \eta_R(s) + \eta_G(s) + \eta_B(s)} \\ \dfrac{G(s) + \eta_G(s)}{S(s) + \eta_R(s) + \eta_G(s) + \eta_B(s)} \end{bmatrix}$$

$$\sim N\left(\begin{pmatrix} \hat{r}(s) \\ g(s) \end{pmatrix}, \sum_{r,g}\right)$$

The detailed expression of the pixel-wise covariance matrix $\Sigma_{r,g}$ is presented at M. Grieffenhagen, V. Ramesh, D. Domaniciu and H. Niemann, "Statistical Modeling and Performance Characterization of a Real-Time Dual Camera Surveillance System," IEEE Conference on Computer Vision and Pattern Recognition, 2000.

See Wikipedia:

http://en.wikipedia.org/wiki/Laplace_transform

Formal definition

The Laplace transform of a function *f*(*t*), defined for all real numbers $t \geq 0$, is the function *F*(*s*), defined by:

$$F(s) = \mathcal{L}\left\{f(t)\right\} = \int_{0^-}^{\infty} e^{-st} f(t)\, dt.$$

The lower limit of 0^- is short notation to mean $\lim_{\epsilon \to +0} -\epsilon$ and assures the inclusion of the entire Dirac delta function $\delta(t)$ at 0 if there is such an impulse in $f(t)$ at 0.

The parameter s is in general complex:

$$s = \sigma + i\omega$$

This integral transform has a number of properties that make it useful for analyzing linear dynamical systems. The most significant advantage is that differentiation and integration become multiplication and division, respectively, by s. (This is similar to the way that logarithms change an operation of multiplication of numbers to addition of their logarithms.) This changes integral equations and differential equations to polynomial equations, which are much easier to solve. Once solved, use of the inverse Laplace transform reverts back to the time domain.

Therefore, it is respectfully submitted that Paragios et al. do not relate to the modeling of images and the processing of image models, but rather the modeling of image sequences (video streams) and a time-based processing of video stream models. (While the present invention may be applied to time sequences of images, the claimed invention requires that a respective image be modeled as a reliable lattice, a feature absent from Paragios et al.). Therefore, while the digitized video streams of Paragios et al. certainly "comprise" image data (Fig. 1, column 4, lines 37-51), the fundamental differences between processing a video stream to perform an object analysis using time differences between frames in a manner tolerant to illumination changes (e.g., shadows) cannot be confused with the analysis of method capable of analyzing a single original image to identify shadows, and the construction of a reliable lattice of an "image" is distinct from the creation of a lattice (see below) representing video information and reflecting its dynamic characteristics.

The examiner indicates that Paragios et al. employ a “reliable lattice” as prescribed by the present claims. A “reliable lattice”, as disclosed in the specification, has probabilities defined on nodes and links (Page 5, line 24-page 6, line 17):

> To overcome the above-identified second and third problems, the inventive method provides a two-level shadow detection algorithm. At the pixel level, the image is modeled as a reliable lattice (RL). The lattice reliability is defined by both node reliabilities and link reliabilities. The inventors have determined that shadow detection can be achieved by finding the RL having the maximum lattice reliability. At the region level, application oriented procedures which remove most possible false detected regions are applied. Since shadow detection can be considered as a special case of image segmentation, the relationship between the RL model and an MRF model such as that taught by Charles A. Bournan, "Markov Random Fields and Stochastic Image Models", Tutorial presented at ICIP 1995 is also developed. MRF models are known to be one of the most popular models for image segmentation. For this reason, their use in shadow detection is important and also allows for possibility of extending the methods of the present invention into more general image segmentation areas. The relationships between RLs and MRFs are developed hereinbelow.

In contrast, Paragios et al. provide a “lattice” as part of the MRF, but not a “reliable lattice” (Col. 5, lines 13-44):

> The change detection/segmentation map 115 is preferably obtained using a Markov Random Field (MRF)-based approach where information from difference sources is combined. Two different motion detection models are proposed. The first is based on the analysis of the difference frame between the observed frame and the most probable background reference state using a mixture model of Laplacian distributed components. The components of the distribution include the samples corresponding to the static background and the moving objects. The second model is intensity-based and has two sub-components: one that stands for the expected background intensity properties (color is assumed) and one that stands for the same properties in a normalized color space. This information is combined within the context of MRFs with some spatial constraints to provide the final motion detection map where local dependencies are used to ensure its regularity and smoothness. The defined objective function is implemented in a multi-scale framework that decreases the computational cost and the risk of convergence to a local minimum. Finally, two fast deterministic relaxation algorithms (ICM, HCF) are used for its minimization.
>
> I. Markov Random Fields
> A general MRF-based framework assumes:
> A finite 2D lattice $S=\{s_i\}$,

> A set of labels L={1ᵢi’ [0,N]}
> A set of observations I={I(s); s’ S}
> And, a neighborhood graph G={g_i, I’ [0,M]} that defines interactions (graph edges) between the pixels (graph sites) of the finite 2D lattice.

Likewise, Paragios et al. nowhere disclose determining a relationship of an RL model of an image with an MRF model, although a type of MRF model is discussed (Col. 3, lines 13-28):

> A video analysis method according to the present invention decomposes the video analysis problem into two steps. Initially, a change detection algorithm is used to distinguish a background scene from a foreground. This may be done using a discontinuity-preserving Markov Random Field-based approach where information from different sources (background subtraction, intensity modeling) is combined with spatial constraints to provide a smooth motion detection map. Then, the obtained change detection map is combined with geometric weights to estimate a measure of congestion of the observed area (e.g. the subway platform). The geometric weights are estimated by a geometry module that takes into account the perspective of the camera. The weights are used to obtain an approximate translation invariant measure for crowding as people move towards or away from the camera.

Paragios et al. state at Col. 8, lines 38-63 that the image space may be segmented, but it is not at all clear that the technique employs “region level verification” as provided by the present claims:

> According to an aspect of the present invention, there are preferably two different approaches to implementing state-dependent classification of image pixels. For example, it is to be appreciated that the architecture of the state model can be fixed in some systems, or adapted to an image sequence in other systems. The former approach involves a fixed design of the network, in which a user-defined, fixed state model is used. In this approach, a user selects K regions in an image based on the context of the image. For example, in an image of a train stop scene, the image may be divided into separate regions corresponding to the train tracks, waiting area for pedestrians, and ceiling area. The number of states Q.sub.k in each region K is defined based on a number of actors n.sub.k present in a region K (K=1, 2, . . . K) and a number of states s_l for each agent (class) l (l=1,2, . . . n_k).
>
> For example, in a train track area, three states may be defined corresponding to: having no train present, a train which is stationary, and a train that is moving. A

> default implementation preferably uses a fully connected Markov chain for each region K. A-priori knowledge about the scene can be used to modify the links in the network. For example, in the above example, certain transitions in state are impossible (i.e., instantaneous transitions from a stationary train to having no train may be zero).

Finally, it is not clear that Paragios et al. **identify** shadow regions in the original image. For example, Paragios et al. state (Col. 4, lines 52-63, Col. 9, lines 42-57):

> Next, for each input frame to be processed the following procedure is preferably followed. In a detection step 109, a change detection map 115 is obtained using, for example, a Markov Random Field based approach in which information from a statistical modeling 111 is combined with spatial constraints 113 and compared with each current input frame from input 101. Thus, the background model 103 is compared with incoming video data to evaluate/detect where change in the images has occurred. In addition, the use of the Markov Random Field framework establishes coherence of the various sources of information in the resulting change detection/segmentation map.
>
> FIG. 3B is an exemplary schematic illustration of the method of splitting a node in a multi-state system for growing a Markov network to find an effective number of states according to an aspect of the present invention. A local model 315 demonstrating multi-modality is split (in accordance with step 311) into multiple nodes 317 and 319. Each of the multiple nodes 317 and 319 is assigned to a new state, thus resulting, for example, in a two-state model here. It is to be noted that the above algorithms used labeled data and fixed regions.
>
> II. c) Invariant Normalized Color Module
>
> Although the color module captures the background intensity properties, it is very sensitive to global illumination changes (e.g. the arrival of a train affects the observed intensities of the platform next to the train line) as well as shadows.

From these sections it is clear that Paragios et al. are responsive to dynamically changing illumination, which may include shadows, but are not limited to shadows per se; while a static shadow present in a series of frames will not trigger a response. Paragios et al. do not specifically target shadows for detection, and provide no means for distinguishing shadows from other causes of illumination changes. Thus, Paragios et al.

are both overinclusive and underinclusive with respect to the presently claimed identification of shadows, and thus fail to teach or suggest the claim element.

Thus, a fundamental difference between the present application and Paragios et al. is apparent: The present application seeks to model shadow behavior based on physical principles, behind a projection of light interacting with a non-transparent object, and analyze an image to extract these features; Paragios seeks to provide a system which is tolerant to changes in illumination (shadows being an example of an illumination change which does not represent an object of interest) while sensitive to desired objects. (Col. 2, lines 56-61):

> Accordingly, an efficient and accurate real-time video analysis technique for identifying events of interest, and particularly, events of interest in high-traffic video streams, which does not suffer from locality and which can handle deformations and global illumination changes, is highly desirable.

Therefore, it is believed that the present claims are clearly distinguished.

Thus, applicants have distinguished Paragios et al. on multiple grounds.

Applicants thus traverse the Examiner's analysis and rejection of claim 2. Claim 1 of Paragios et al. is expressly limited to video analysis. Col. 4, line 63-Col. 5 line 3, while discussing a single video frame, analyze this frame within the context of its sequence, and therefore fail to satisfy the claim limitations:

> The change detection map 115 is then combined with the geometry information 107 (step 117) to estimate congestion of the observed input frame (step 119). Then, using the change detection/segmentation map 115 combined with the current video frame (i.e the observations), the background model 103 is updated mainly, for example, for pixels in the current frame that are labeled as static pixels in an updating step 121. The process 100 is then repeated for a next input frame.

In any case, the method disclosed in Paragios et al. is inoperative and not enabled to detect shadow regions in an original image which is a single, static image.

With respect to claim 3, the examiner cites Paragios et al., Col. 4, lines 26-26, which states:

> The subway video analysis application has requirements such as real-time processing on compressed video streams, low cost, camera viewpoint, etc. Moreover, the illumination conditions are characterized by near static situations mixed with occasional sudden changes due to change in platform state (e.g., ambient illumination changes due to train arrival/departure in the scene). The task considered in the present invention involves determination of the congestion factor in subway platforms. Congestion is defined as a prolonged temporal event wherein a given percentage of the platform is crowded for a user-defined period of time.

In fact, this disclosure states nothing about single point illumination, and it is believed well known in the art that subway platforms are illuminated by a plurality of sources, and best practices in the design of subway platforms would seek to minimize shadowing to improve passenger safety.

With respect to claim 4, it is respectfully submitted that Paragios et al. do NOT refer to the sun when they employ the phrase "global illumination changes"; in fact, they appear to be referring to changes in illumination of the entire frame.

With respect to claim 5, Paragios et al. do not discuss aerial photography, and no analogy to the claims referenced is observed.

With respect to claims 6-8, as stated above, Paragios et al. do not teach or suggest use of reliable lattices, and therefore do not teach or suggest the substep of modeling an initial RL and/or updating the model and/or iteratively updating the model.

With respect to claims 10-11 Paragios et al. do not discuss determining a reliability or maximum reliability of a reliable lattice, and no analogy to the claims referenced is observed.

With respect to claim 13, the shadow must be detected in accordance with the method of claim 1, and then a "false shadow" removed. This technique is neither taught nor suggested in the references, and it is respectfully submitted that no prima facie case of obviousness is presented. As noted on p. 3, a predicted image is required by Jaynes for each view, which is not provided by the method of Paragios et al.

With respect to claim 14, the examiner equates a normalized RGB space and a normalized LogRGB space, thus trivializing the claim and ignoring a particular claim element.

With respect to claim 15, a region level verification is performed in addition to step (d) of claim 1, which is neither taught nor suggested in the references.

With respect to claim 16, domain knowledge is exploited to perform the region level verification, also not taught or suggested in the references.

Claims 9 and 12 are allowed.

Claims 17-20 are new. It is believed that new independent claim 20 expresses the same inventive concept as claim 1, and that no restriction is appropriate.

Respectfully submitted,

Steven M. Hoffberg
Reg. No. 33,511

MILDE & HOFFBERG, LLP
10 Bank Street - Suite 460
White Plains, NY 10606
(914) 949-3100